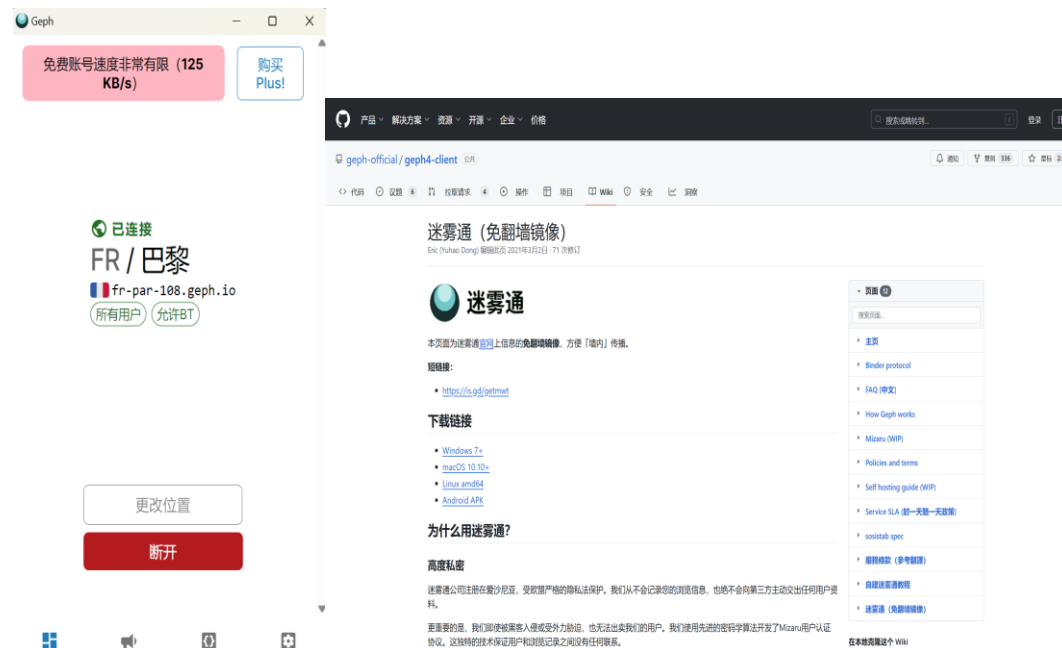


准备好梯子（还有足够的耐心）十几个 GB 的存贮

[迷雾通（免翻墙镜像）](#) · [geph-official/geph4-client Wiki](#) · [GitHub](#)（一个免费的梯子）

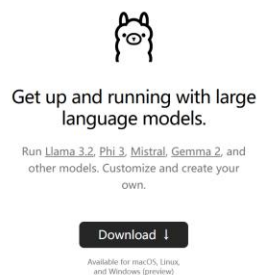


B 站视频：【Llama 3.2 一键本地部署！支持视觉、图片识别功能！Meta 最强开源大模型，100% 保证成功！！ | 零度解说】

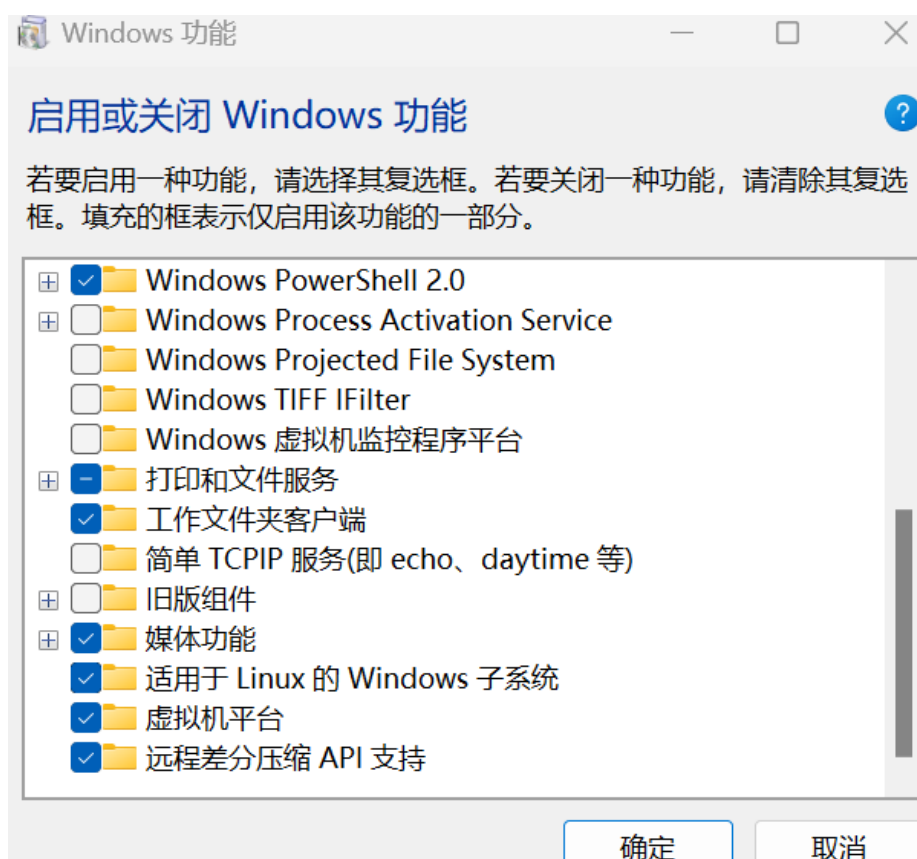
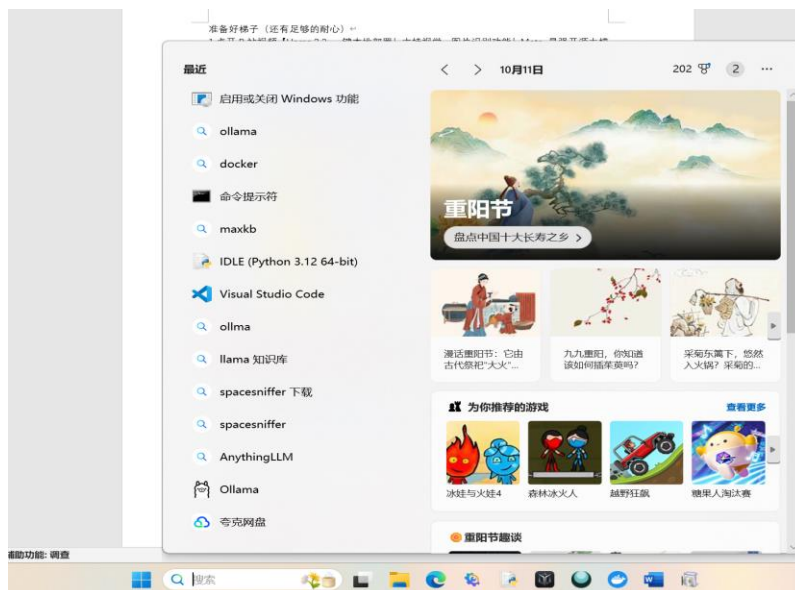
https://www.bilibili.com/video/BV1bkx1eAEuc/?share_source=copy_web&vd_source=6bf8eb1c6205929a58c7bccc8626b8a6

<https://www.freedidi.com/14712.html> (llama 3.2 3B) (安装步骤 ollama)

#1 安装 ollama(<https://ollama.com/>)进去 download 就行，看你的操作系统选版本



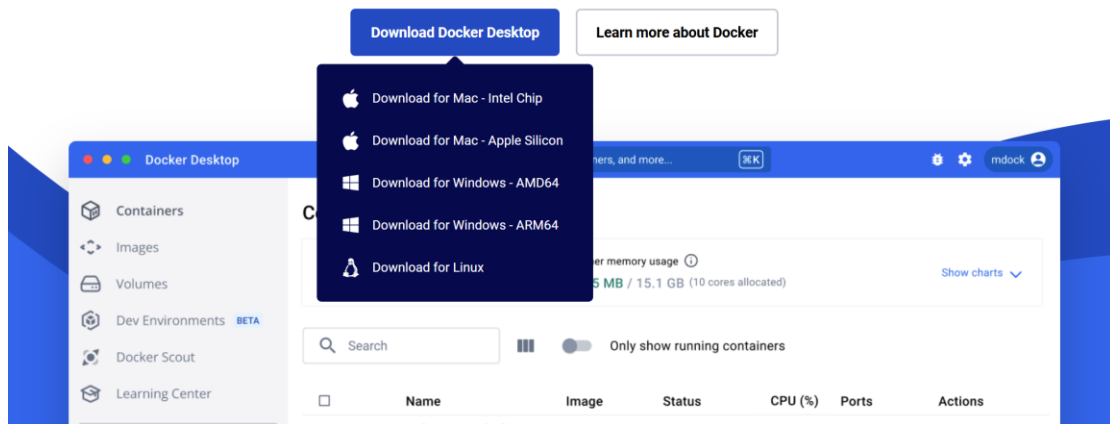
#打开 WINDOWS WSL 功能 搜索启用关闭 WINDOWS 功能，勾选上适用于 LINUX 的 WINDOWS 子系统——然后重启



#安装 DOCKER 桌面版 (<https://www.docker.com/>) 下载，根据自己版本选（进去记得挂梯子）安装好——重启——重启之后点 FINISH 就行（之后的 DOCKER 账号可登录也可以不登录）

Run anywhere.

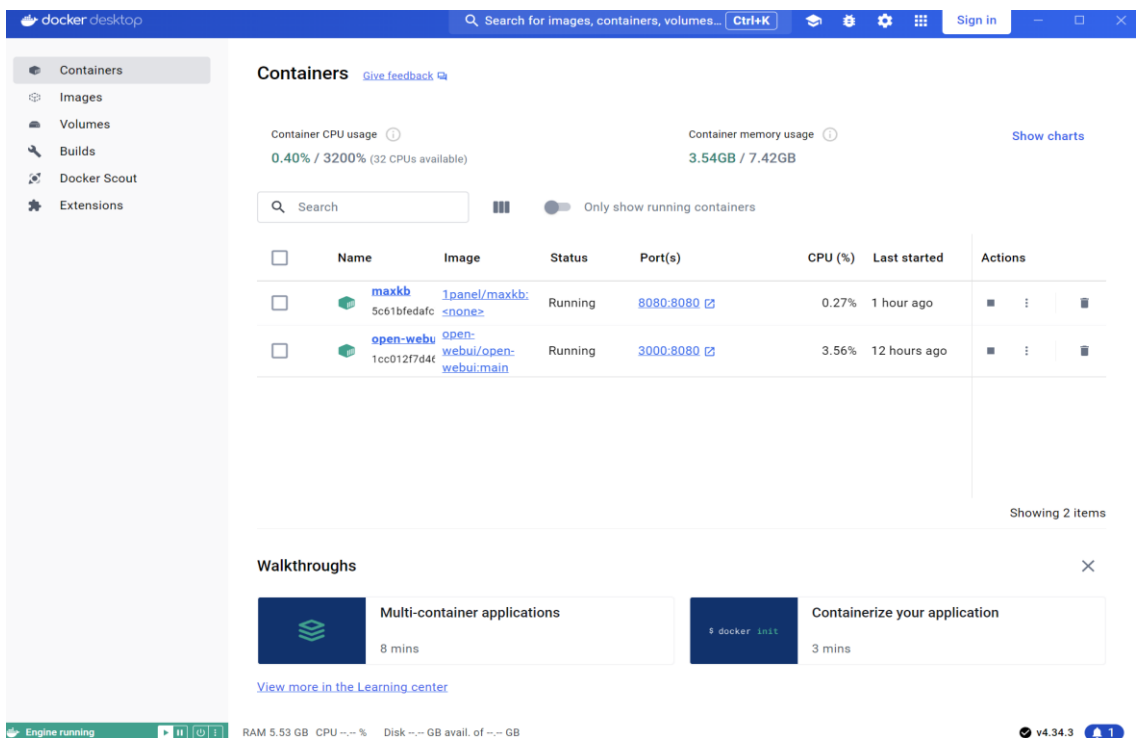
Build with the #1 most-used developer tool



【AMD和ARM】

1、上面也提到了，amd是一家CPU制造商，它们家的CPU和Intel一样，都是X86架构。中文名（超威）超微半导体，是除了英特尔以外最大的x86架构微处理供应商（即：美国超威半导体公司），也是除了英伟达以外仅有的独立图形处理供应商。

2、ARM（英文为Advanced RISC Machine，或Acorn RISC Machine），应该这么说，arm和x86是一类，也是CPU的一种架构；非常适用于移动通信这种低成本，高性能，低耗电的领域。ARM的公司为安谋控股（ARM Holding plc），又称为ARM公司，总部位于剑桥的设计与软件公司，现在已被日本的软银公司收购，后者同时也是国内互联网巨头阿里巴巴的最大股东。



#本地安装 WEBUI（docker run -d -p 3000:8080 --add-host=host.docker.internal:host-gateway -v open-webui:/app/backend/data --name open-webui --restart always ghcr.io/open-webui/open-webui:main）这个命令直接输 CMD 里面 ENTER，如果不行，出

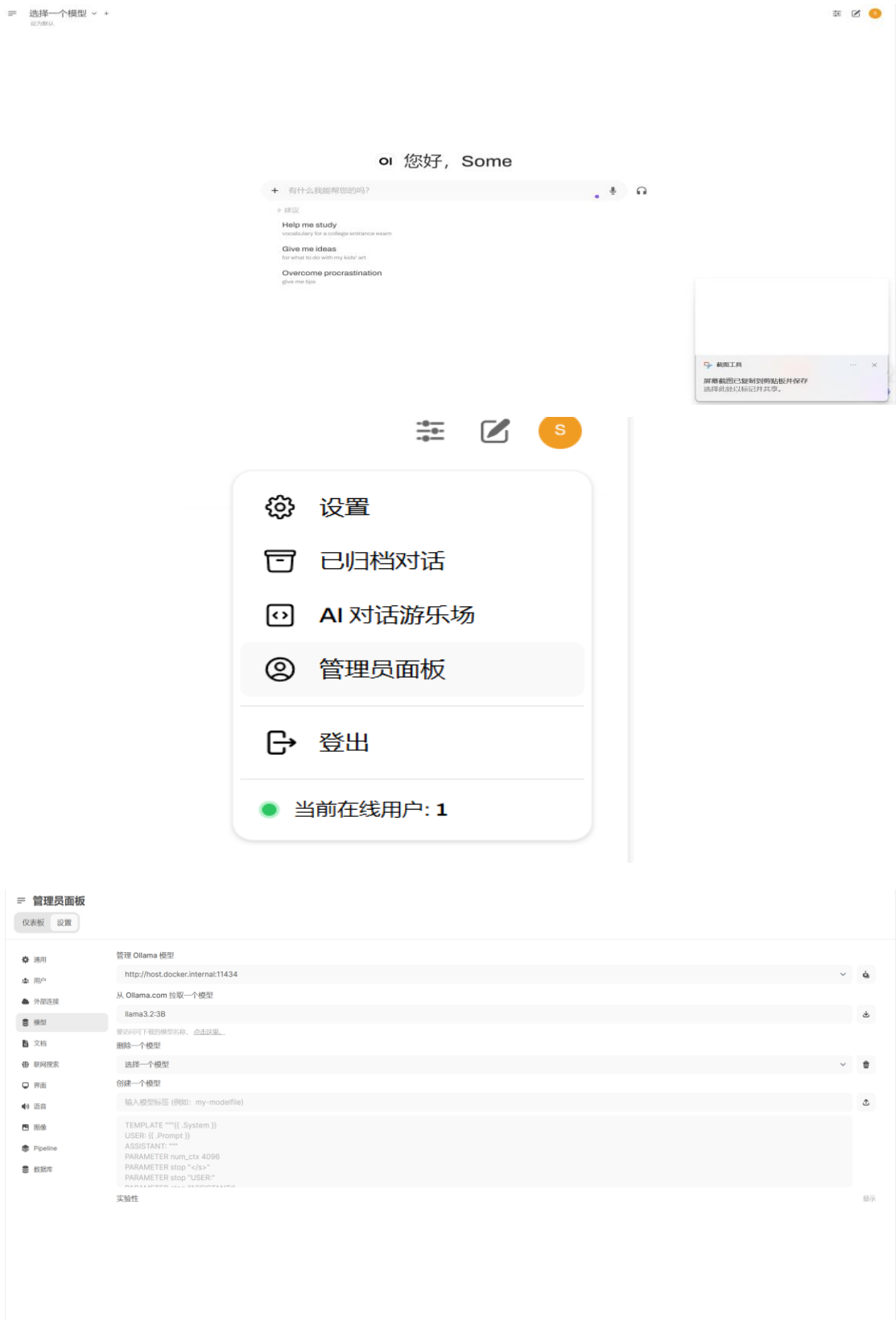
现报错什么的（重启一下或者等一会）（这个地方可能需要科学上网，不科学上网行不行我没试过）



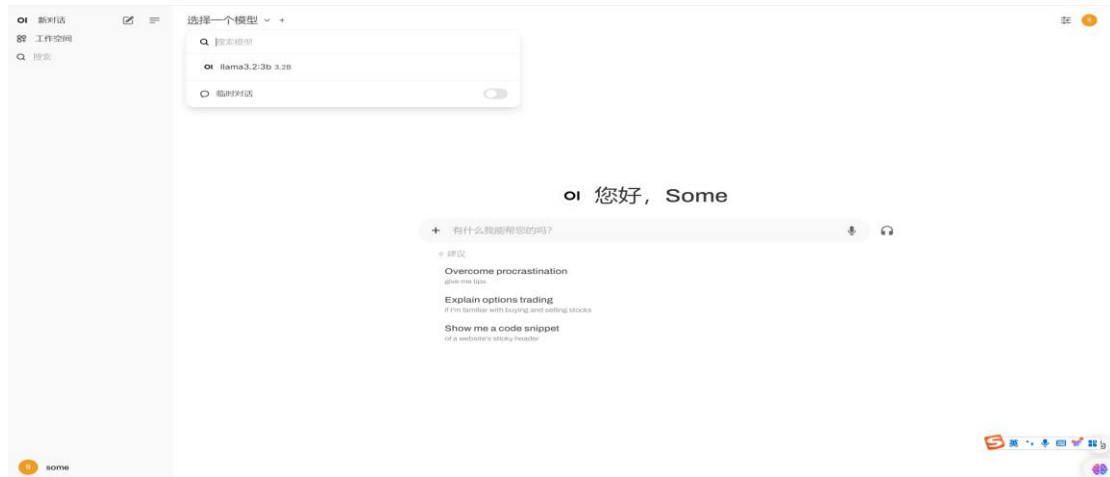
#5 安装完成通过本地地址：http://127.0.0.1:3000 进行访问，（账号什么的注册好）



#左上角或者右上角找到有你名字首字母的那个头像，点击设置，进入管理员面板，找到模型，输入模型名称（不知道名字自己点下方查，但是由于奇奇怪怪的问题不一定能打得开）这里安装 3B（要求显卡显存 6G）的话，在那一行输入（llama3.3:3b）就行，右边下载点一下（可能需要一段时间）



#回到上述的首页，点击新对话，选择模型 llama3.2:3b,然后问问题就行



2.MaxKB 构造知识库

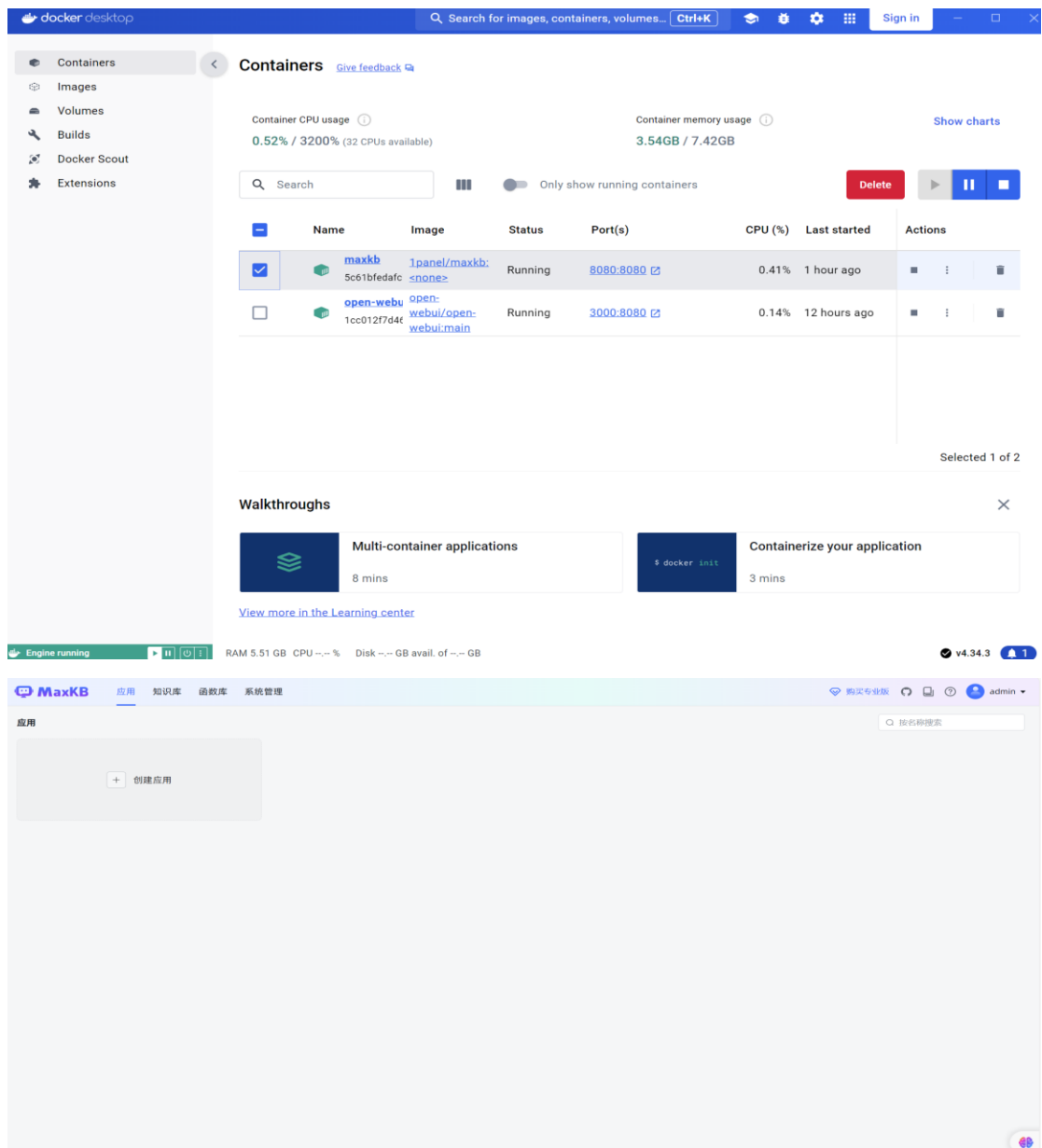
B 站参考视频：【MaxKB 篇：手把手教你打造一个本地安全可靠的知识库】
https://www.bilibili.com/video/BV1MD421T76N/?share_source=copy_web&vd_source=6bf8eb1c6205929a58c7bccc8626b8a6

步骤# <https://github.com/1panel-dev/MaxKB>（记得挂梯子，在里面找到快速开始的那个命令（`docker run -d --name=maxkb --restart=always -p 8080:8080 -v ~/.maxkb:/var/lib/postgresql/data -v ~/.python-packages:/opt/maxkb/app/sandbox/python-packages cr2.fit2cloud.com/1panel/maxkb`）cmd（命令提示符）输进去然后 ENTER（不用挂梯子的）

```
命令提示符
Microsoft Windows [版本 10.0.22631.4317]
(c) Microsoft Corporation. 保留所有权利。

C:\Users\lenovo>docker run -d --name=maxkb --restart=always -p 8080:8080 -v ~/.maxkb:/var/lib/postgresql/data -v ~/.python-packages:/opt/maxkb/app/sandbox/python-packages cr2.fit2cloud.com/1panel/maxkb
```

#完成后打开 DOCKER，找到新增容器，点击右边的 PORT 进入



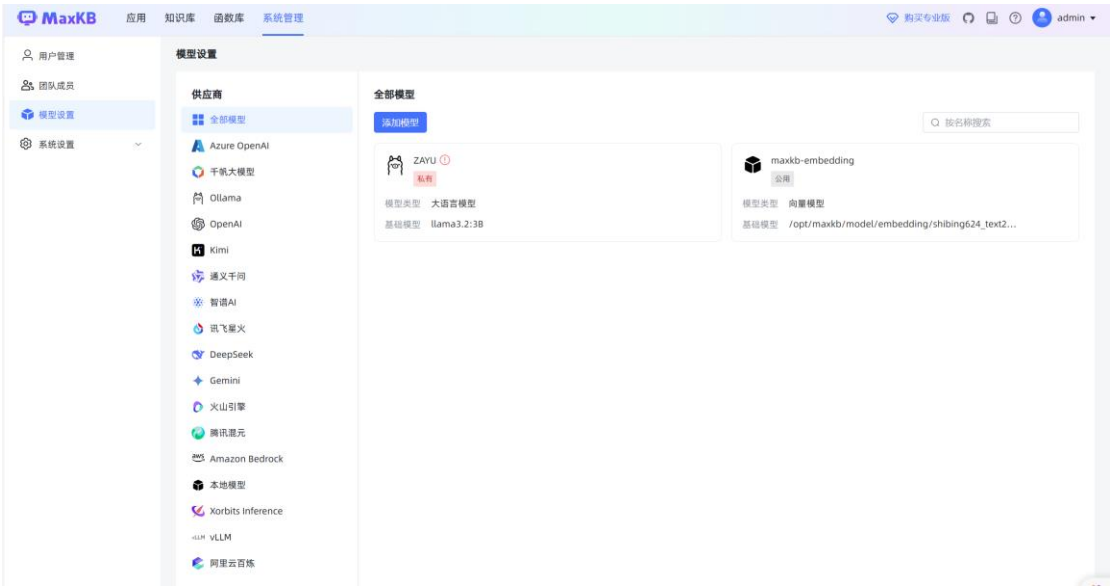
用户名: admin

密码: MaxKB@123..

第一次进入账号密码默认（输上面的，进去会弹出密码设置）



系统设置——模型设置——添加模型——选择 ollama



编辑 Ollama

模型名称 ⓘ *

some jiba caotamd sdwdas 24 / 64

权限

☒ 私有
仅当前用户使用

☐ 公用
所有用户都可使用，不能编辑

模型类型 *

大语言模型

基础模型 列表中未列出的模型，直接输入模型名称，回车即可添加 *

llama3.2:3b

API 域名 *

http://host.docker.internal:11434

API Key *

.....

取消

修改

模型类型，基础模型（也可以填别的，下面有选项，这个 3.2:3B 我还没试过）API 域名如上图填写（http://host.docker.internal:11434） API KEY 模型名称随便写



知识库的创建



应用

——总结来源于人肉智能(手打的)